Association between Time of First Publication and Future Publication Success in Research-Intensive Medical Academicians: A Cross-Sectional Study

By

Kevin R. Riggs

A Master's Paper submitted to the faculty of
the University of North Carolina at Chapel Hill
In partial fulfillment of the requirements for
the degree of Master of Public Health in
the Public Health Leadership Program.

Chapel Hill

July 2009

MANUSCRIPT:

ABSTRACT:

Background/Aims:

Publication during medical school and residency has been associated with higher rates of publication and more citations of one's published research. However, it is unclear whether that association is seen and persists late into the careers of research intensive academicians. The purpose of this paper is to investigate this latter supposition.

Methods:

The authors examined the CVs of corresponding authors of articles from NEJM and JAMA from 2008. Publication records of the authors were extracted from their CVs, and citation records for each author were obtained from the Web of Science. Analysis of covariance regression models were used to examine the association between first publication during medical school, within 5 years of medical school, or beyond 5 years of medical school graduation, and future publication and citation success.

Results:

On adjusted analysis, those researchers who had their first publication before medical school graduation had a greater number of mean publications after medical school and during 2006-2007 (163.6 and 28.4) than those who had their first publication during the first 5 years after medical school graduation (110.7 and 18.8) or more than 5 years after graduation (59.0 and 13.1). Similarly, those who had their first publication before medical school graduation had a greater number of mean citations to their published work since graduation and on publications from 2006-2007 (4634 and 333) than those who had their first publication during the first 5 years after medical school graduation (2936 and 183) or more than 5 years after graduation (1512 and 143).

Conclusions:

Early publishing is associated with higher rates of publications and more citations of published research, and this association is maintained well into a researcher's career.

INTRODUCTION:

Medical students conducting research is common.[1] It has long been argued that participation in research early in one's career is absolutely critical to those who wish to go into academic medicine, and is even beneficial to those who do not.[2] The stated benefits of participating in research as a medical student are that the experience gives students a better ability to critically appraise new scientific discoveries, setting the stage for a lifetime of learning. Another positive benefit is that doing research as a medical student is associated with choosing a career in academic medicine.[3] This second reason has dominated the research recently, given the worry of the past several decades that the clinical investigator is an "endangered species."[4] However, a less articulated, but widely held belief is that participating in research as a medical student teaches the student how to perform scientific research early, which will benefit their future research.

As the amount of medical knowledge expands, however, medical students are expected to learn more in the same amount of time, and there may potentially be less free time available for medical students to conduct research. Some students, especially those who know that they want to go into academia, may be tempted to put off the start of their research career until residency or fellowship, since many programs offer protected time to do research. As such, it is important to know exactly what the benefits of conducting research during medical school are in order for students to make informed decisions as to whether it is worth their time to participate.

Previous studies have demonstrated an association between doing research or publishing in medical school and future academic success.[5, 6] However, these studies have been conducted in academic settings in general, and not specifically among research-intensive medical researchers. Furthermore, it is not clear whether the association attenuates over time.

To that end, we conducted this study to determine whether there is an association with publishing early in one's career and future publication success, and whether that association holds over time.

MATERIALS AND METHODS:

This research was conducted at Duke and determined to be exempt research by the Duke IRB. UNC IRB approved this study as existing or non-research data.

Subjects:

To identify potential researchers, we looked at each issue of the New England Journal of Medicine (NEJM) and JAMA for the year 2008. We contacted corresponding authors from the Original Articles (including Brief Reports) from NEJM and Original Contributions from JAMA who had an MD, MD equivalent (DO, MBBS, MB ChB, etc), or potential MD equivalent (FRACP, FRCP, etc), and did not have another doctoral level research degree (Ph.D., DPH, etc) by email and asked them to send a current *curriculum vitae* (CV).

To compare the group of responders to non-responders, we extracted several factors from the articles, including whether the research was an randomized controlled trial (RCT), whether the mailing address of the author was in the US or not, whether the investigator had another post-bachelor level degree, and whether the researcher was first author, last author, or neither. For authors with more than one article published during this time frame, a single email was sent, using information from the earliest eligible article.

Extraction of Data from CVs:

Data from CVs were extracted independently by two people (KR and ZR), and any discrepancies were resolved by consensus. Data extracted from CVs included year of graduation from medical school, whether medical school was in the US, residency type, other post-bachelor level degrees, number publications published before medical school graduation, number of publications published between 1-5 years after medical school graduation, number of publications published after medical school graduation through the end of 2007, and number of publications published in 2006-2007.

Gender was not self reported, and was thus determined by the extractors. In cases where gender was not able to be determined from the name or from context within the CV, we conducted an online search to find a picture of the researcher, typically from their current institutional profile. There were no cases where consensus on gender of researcher was not able to be reached.

Our intention was to include all written publications that referred to specific works without including abstracts. We did include letters, proceedings of meetings, and online publications where the author was cited by name. We did not include entries that referred to non-written media (DVDs, CDs, etc), works that were only edited, or educational materials where the author was not cited by name. In order to exclude abstracts, we did not include entries under "abstracts" or any header that included abstracts (i.e. "abstracts/letters"). We also excluded entries from "miscellaneous" or "other" unless their contents were further described.

Extraction of Data from the Web of Science:

We searched Web of Science (WoS) for both publications and total citations for the time periods consisting of medical school graduation through 2007 and 2006-2007. Each search was conducted on the Science Citation Index Expanded (SCI-EXPANDED) using the researchers name as it was written in the index article (last name, first initial, and possibly middle initial). For women researchers, we checked the first publication on the CV to determine if there was a name change. In cases where there was a name change, we searched both names.

Author attribution, determining who wrote a specific piece of literature and identifying all of the works of a given individual, is a significant gap in the field of information sciences, and much research is currently being conducted in the field to address it.[7] We chose to use a relatively simple method that maximized specificity (not attributing papers to the author of interest) rather than sensitivity (not missing papers that should have been attributed to the author of interest). Our primary method of disambiguating was to limit searches by institution according to the institutions listed on the researcher's CV.

Each search was also bounded by the years of interest. In cases where the researcher had published before medical school graduation, the search was performed from the year after medical school graduation through the end of 2007. If the researcher did not publish before medical school graduation, the search was bounded by the year of first publication according to the CV. For example, if according to the CV the first article published was 6 years after medical school graduation, then no search was conducted for years 0-5. The reason for this was to ensure that we did not falsely attribute articles to any researcher before their stated first publication.

After conducting the search for publications after medical school through the end of 2007, we examined the number of publications found on WoS compared to the number on their CV. If the number found on WoS was less than 50% of the number found on the CV, we went back to the CV to look for any alternative name spellings that could explain the low yield. If the number found on WoS was greater than 100% of the number found on the CV, we assumed that there may be more than one author by the name searched at one of the institutions searched, and we attempted to exclude any papers that did not belong to the researcher of interest.

Data Analysis:

The four main outcome variables were total publications from medical school graduation through 2007 as determined from the CV and citations on those publications, and publications during 2006-2007 as determined from the CV and citations on those publications. The main exposure variable was time of first publication, as defined in three categories: "early publishers" (first publication on CV was on or before the year of medical school graduation), "middle publishers" (first publication on CV was between 1-5 years after medical school graduation), and "late publishers" (first publication on CV was >5 years after medical school graduation).

The analysis of the four main outcomes involved a formal examination of potential confounders, including association of confounders with the main exposure (Table 2) and bivariate association of confounders with each of the four main outcomes. Linearity between continuous variables and outcome measures as well as collinearity between each of the categorical variables were evaluated prior to modeling.

For each of the four main outcome measures, a multiple linear regression analysis of covariance (ANCOVA) model was used to estimate the mean number of publications or citations, adjusted for covariates. Because time of first publication was divided into three categories, indicator variables were created, with early publishers as the reference category. For the publications from medical school graduation through 2007 and citations on those publications, models were adjusted for years since medical school graduation (as of the end of 2007), gender, whether the participant went to medical school in the US or not, and presence of an additional graduate level degree (none, MPH, or other masters degree). For the publications during 2006-2007 and citations on those publications, the models were adjusted for the same variables, except that stage of career (a four-category nominal variable) was used, rather than years since medical school as a continuous variable. Ideally, we would have liked to adjust for residency type in all of the analyses, but the large percentage of participants who did internal medicine (including subspecialties) meant that the remaining residency types had too few entries for a meaningful regression model.

RESULTS:

Of the 199 researchers emailed, 102 sent us a CV. Characteristics of responders and non-responders are listed in Table 1. Those who responded were more likely to have a correspondence address in the US ($p=0.04$). No other differences were statistically significant.

Of the 102 CVs sent, 83 were included in the final analysis. Reasons for exclusion included researcher had a PhD (6), CV did not contain a publication list (4), publications on CV were marked as "selected publications" (3), CV was not updated through the end of 2007 (2), insufficient information on CV (3), and CV was sent as an un-openable file (1). In the case of selected publication list or no publication list, we would have been unable to accurately categorize researchers as early, middle, or late publishers, which was a critical first step. We excluded those with insufficient information on the CV because we were either unable to determine when they graduated from medical school, or had no way of applying our author disambiguation strategy because they did not indicate any institutions with which they had been associated. Finally, we excluded those CVs that were not updated because we had expected that all CVs would have been updated through that point, and we had no *a priori* plan to deal with cases where they were not.

Characteristics of early, middle, and late publishers who were included in the final analysis are listed in Table 2. Notably, the early publishers were more likely to be male and were about four years "younger" than the middle or late publishers.

Early publishers had a mean number of publications before graduation from medical school of 3.2 (range 1-25), and 5.3 during the first 5 years after medical school (range 0-33). Middle publishers had a mean number of publications during the first 5 years after medical school of 3.2 (range 1-12) By definition, middle publishers had no publications before medical school graduation, and late publishers had no publications before medical school graduation or during first 5 years after medical school graduation.

The results of unadjusted and adjusted analyses are displayed in Table 3. In all four outcomes, after adjusting for covariates, early publishers had 47-82% more publications/citations than the middle publishers and 116-206% more publications/citations than the late publishers. For the publications after medical school graduation through 2007 and citations on these publications, and publications in 2006-2007, each of these differences were statistically significant. For the citations on publications in 2006-2007, these differences were only marginally significant ($p \leq 0.083$).

DISCUSSION:

This analysis has demonstrated that an association exists between early publishing and future publication success among a group of research-intensive medical academicians. Furthermore, examination of the most recent two year period of publication demonstrates that the association is maintained well into one's career.

The difference in publication rates extending late into an academician's career suggests that an early research project which leads to publication truly has a beneficial effect of significant magnitude on one's future career. However another possibility is that there are still some uncontrolled confounding effects at play. In a cohort study, where those who did research in medical school are compared to those who did not, significant differences in personal characteristics, such as motivation and scientific curiosity, are likely to bias the results. Yet, by sampling researchers as we did, under the assumption that to be the "lead" author on an original study published in a high impact journal requires at least a significant degree of research skill, personal characteristics such as motivation or scientific curiosity, are unlikely to be causing significant confounding.

There were several limitations of this study. First, the small sample size and stringent inclusion criteria of this study makes it difficult to generalize the results to all medical researchers. However, the difference was profound enough that even with our small sample size, the results were significant. Second, our assumption that each researcher identified would be a "research-intensive" academician may also not have been true. We did not ask the researchers to self report what percentage of their time was spent doing research, but it appeared from the CVs that the vast majority held academic positions, and the results themselves demonstrate that this group consisted of prolific publishers. While perhaps not true for each individual, we feel that the group as a whole could be considered "research-intensive."

Third, we essentially used publication during medical school as a marker of having conducted research during medical school. Ideally, research should be undertaken with the goal of publishing findings, but this is certainly not always the case. Given the potential difference in degree of scope of research projects undertaken by medical students (several weeks vs. a summer vs. a year or more), we feel that publication of research may actually be a better measure of a meaningful research project than self-reported experience. Fourth, we were unable to control for residency type due to the small number of participants from several of the major specialties. However, the fairly even distribution of residency types among the three groups makes it unlikely that it would have been a major confounder. Finally, the cross-sectional survey design makes it difficult to assess causality.

The value of medical student research can be viewed from several vantage points. Competitive residency programs may be interested in accepting those who are likely to go on to become good researchers. In that case, the association between early research and future success may be sufficient, even if it is only a "marker" of future success, and not a contributing factor. For medical schools, who invest money in research opportunities, and students, who have to invest additional time during an already busy time and may potentially delay their graduation, conducting research may not be in their best interest if it is only a marker. While this study was not designed to definitively determine the effect that early research has on future success, the results contribute additional evidence that there may be a true beneficial effect. More work is warranted to assess the true effect that early research has on future research success. However, students who plan to enter a career in academia may be well advised to act on this strong association alone.

FIGURES AND TABLES:

Table 1. Comparison between responders and non-responders.

Characteristic	Responders percent (n = 102)	Non-Responders Percent (n = 97)
Address in US*	74	60
RCT	44	42
First Author	72	71
Last Author	23	24
Non-Phd post graduate degree		
MPH	12	11
Other	14	10

§ Significance tests for comparisons between Responders and Non-Responders based on Pearson's chi-square test for categorical characteristics
* p<0.05

Table 2. Baseline characteristics of respondents.

Characteristic	Overall Mean or % (n = 83)	Early Publishers Mean or % (n = 24)	Middle Publishers Mean or % (n = 42)	Late Publishers Mean or % (n = 17)
Male (%)*	76	92	74	59
Medical School in US	69	83	64	59
Years Since MS graduation (mean)	20.7	17.8	21.6	22.8
0-8	11	25	7	0
8-16	27	21	26	35
16-24	27	25	29	24
24+	36	29	38	41
Non-PhD post graduate degrees				
MPH	16	13	14	24
Other	29	25	29	35
Specialty				
Internal Medicine	57	63	62	35
Surgery	6	4	10	0
Pediatrics	12	8	10	24
Obstetrics and Gynecology	6	4	5	12
Other	19	21	14	29

§ Significance tests for comparisons between early publishers and late publishers based on 2-sample t-test for continuous characteristics and Pearson's chi-square test for categorical characteristics
* p<0.05

Table 3. Results.

Results	Early Publishers Mean (n = 24)	Middle Publishers Mean (n=42)	Late Publishers Mean (n = 17)	p value
Unadjusted				
Publications after medical school	136.9	118.6	77.3	0.229
Citations on publications after medical school	3925.3	3176.9	1919.0	0.234
Publication in 2006-2007	25.0	19.9	15.3	0.168
Citations on publications in 2006-2007	298.5	195.1	160.6	0.168
Adjusted				
Publications after medical school*	163.6	110.7	59.0	<0.001 ŧ
Citations on publications after medical school*	4634.2	2936.4	1512.3	0.002ƚ
Publication in 2006-2007§	28.4	18.8	13.1	0.011 †
Citations on publications in 2006-2007§	332.8	182.7	142.9	0.145 ‡

* Adjusted for years since med school graduation, gender, location of med school (US or not US), and additional degrees (none, MPH, or other masters)

§ Adjusted for stage of career, gender, location of med school (US or not US), and additional degrees (none, MPH, or other masters)

ŧ Overall p value (2df); For individual comparisons: Early vs. Middle, p=0.008; Early vs. Late, p<0.001; Middle vs. Late, p=0.017.

ƚ Overall p value (2df); For individual comparisons: Early vs. Middle, p=0.014; Early vs. Late, p=0.001; Middle vs. Late, p=0.054.

† Overall p value (2df); For individual comparisons: Early vs. Middle, p=0.019; Early vs. Late, p=0.004; Middle vs. Late, p=0.179.

‡Overall p value (2df); For individual comparisons: Early vs. Middle, p=0.077; Early vs. Late, p=0.083; Middle vs. Late, p=0.657.

NON-MANUSCRIPT ADDENDA, UNC MPH THESIS:

Research Question:

Among a selected cross-section of successful medical researchers, is first publication during medical school, compared to first publication during first five years after medical school or greater than five years after medical school, associated with higher levels of publication and citation to their work, both for their entire career and for a recent two year period?

Hypothesis:

My hypothesis is that early publication will be associated with higher levels of publication and citations early in their career, but that eventually, those who had their first publication later will "catch up", and their publication rate and citation rate will be similar to the early publishers. Thus, the early publishers will have more publications and citations for their career, but there will be no difference for the recent two year period. (Essentially, that this cohort of successful academicians will currently be doing a similar quantity and quality of work, regardless of whether there was early publication or not.)

Addendum to Introduction (See page 3):

Systematic Review – Association between Early Publishing and Future Success:

The question being asked in the current study is essentially whether early publishing is associated with future academic success. Therefore, we decided to do a systematic review on the same question to determine what studies have already addressed this topic. Our focused question, in PICO format, was "Among medical physicians, is conducting research or publishing either before medical school graduation or during residency, compared to those medical physicians who did not, associated with higher future rates of publication, citations to published work, academic promotion, or attainments of grants?"

We searched the MEDLINE/PubMed database (January 1975 to June 2009) using the following search terms: "education, medical," "research," "publishing," "career mobility," and “achievement.” For a detailed list of systematic search terms and limits, please refer to Table 13 in the Appendix.

We reviewed abstracts of articles in peer-reviewed journals. Because of the small number of studies addressing our study question, we considered all research designs. However, we did exclude commentaries, letters, or other articles which did not describe original research. We also excluded articles that were primarily concerned with MD-PhD students (although we did not exclude articles simply because they included some MD-PhDs). Using the described criteria, we identified five studies that addressed our research question. We rated the internal and external validity as described by Harris et al,[8] and the evidence table is shown in Table 4.

Brancati et al [5] followed a cohort of graduates from a single institution, and looked at early factors that affected future academic success. They found that research experience in medical school was independently associated with having chosen an academic career (RR=1.8, $p<0.001$). Using multiple regression, they also found that among those who chose academic careers, research experience in medical school was independently associated with both higher academic rank (RR=3.11, $p=0.0001$) and more citations to their published work (ratio=2.42, $p=0.0008$). Membership in AOA and ranking in the top third of their class was also associated with higher academic rank and more citations.

Dorsey et al [6] conducted a retrospective cohort of neurology residents from a single institution. They used five metrics to assess academic outcomes, and assessed what early factors were related to those outcomes. Using regression models, they found very few factors that were significantly related to future success. However, publishing before ($R2=0.11$, $p=0.01$) and during residency ($R2=0.26$, $p<0.001$) were associated with higher publication scores (defined as the number of total publications per year after residency completion divided by the number of years since completion).

Reinders et al[9] describe a Dutch study in which a cohort of physicians who had been admitted to medical school in 1982 or 1983 and who were surveyed in 1997 regarding research during medical school. They then did a Medline search in 2002 to determine the number of articles published after graduation. They found that those who had participated in research during medical school published more after graduation (average 4 articles) than those who did not (average 1 article).

Evered et al[10] performed cross-section study of medically qualified professors and readers in medical faculties in the UK. They categorized groups as being either graduates of Oxford or Cambridge, having an intercalated BSc degree (both of these groups supposedly having very high rates of research during medical school), or graduates from other schools without an intercalated degree, and examined publication rates, citation counts, and grants received by each group. They performed citation analysis on a small subset of the group and found that the academics with intercalated degrees (mean 8.04) and from Cambridge or Oxford (mean 7.63) had more citations per paper than those from other schools without an intercalated degree went on to have more citations per publication (mean 4.16). Several tables of results are listed, but it is not possible to determine whether the majority of their findings were statistically significant. They concluded that "research training or experience, or both, as an undergraduate has a positive influence on career intentions and subsequent research performance".

Segal et al[11] performed a sort of natural experiment when they compared the postgraduate medical experience of graduates of one medical school that required research during medical school (Pennsylvania State University) and two similar medical schools that did not require research (University of Massachusetts and University of Connecticut). While they were primarily examining the effect that research had on choosing to enter an academic career, they also found that those with medical school research experience were more likely to publish the results of postgraduate research than those who did not have medical school research experience (49% vs. 32%, p<0.001). However, they also found that those who had participated in research in medical school were more likely to have participated in postgraduate research (both clinical and basic science, p<0.001), and there was no attempt to control for this factor in measuring publishing rates.

Dyrbye et al[12] conducted a retrospective cohort study of graduates from Mayo Medical School (MMS) from 1976-2003. Students at MMS are required to complete a research experience during medical school. The length of "protected" research time has decreased from 21 weeks to 13 weeks over the course of the time period, but students are able to continue their research on their own time or during elective research time. The study had several aims, but they found that those who published a research report related to their required research (0.5 vs. 0.3, p<0.0001), published an abstract related their required research (0.5 vs. 0.4, p=0.04), or presented their research at a meeting (0.5 vs. 0.3, p=0.001) had more unrelated publications during the first three years after medical school than those who did not.

Chusid et al[13] conducted a retrospective cohort study of the 79 surviving members of the Yale School of Medicine class of 1970. Students in that year were required to complete an original thesis as a prerequisite to graduation. Medline was searched from 1970-1990 to determine whether the medical school thesis was published, and also the total number of publications for each class member over that time period. They found that those who had published their thesis had more publications than those who did not (22.1 vs 14.4, p=0.005). No attempt was made to determine what proportion of the class had chosen an academic career.

Lessin et al[14] conducted a retrospective cohort study of 248 physicians who began a pediatric surgery residency (PSR) between 1979-1992. PSR occurs after general surgery residency (GSR), and the authors were examining whether research and publication during GSR predicted publication during and after PSR. Among the 200 who responded to the survey, those who had conducted laboratory research during their GSR published more laboratory publications (3.7 vs 0.7, p<0.00001) and clinical publications (3.7 vs. 2.5, p<0.05) than those who did not conduct laboratory research, but there was no significant differences in the number of publications during PSR or after PSR. They also found that those who published during GSR had more publications during PSR (4.5 vs 2.1, p-value not reported) and total publications after GSR (11.9 vs. 5.1 p-value not reported) than those who did not publish during GSR. It is not clear in either of these analyses how long the time after PSR was, or that it was equal among the groups.

Smith et al[15] conducted a retrospective study to evaluate the effect of the American Pediatric Society-Society for Pediatric Research Medical Student Research Program. They compared those medical students who had been awarded the 8-12 week research scholarship from 1991-2000 to those who had applied and not received the award. Using a PubMed search (done in 2001), they found that participants were 79% more likely to have published than nonparticipants ($p<0.0005$), and that the number of publications per person were higher for participants than non-participants ($p=0.007$, but numbers not able to be determined from paper).

Patterson et al[16] conducted a retrospective cohort study of 78 radiology residents who completed their residency at a single institution between 1990-2000. They were examining whether research performance from undergraduate or medical school was associated with research productivity during and immediately following residency. They found that there was no significant difference between the number of papers published by those who were "previously productive" during medical school (which was defined as any posters, articles, book chapters, abstracts for which the student was listed as a coauthor, or research projects in which the student was involved in the absence of any authorship) and those who were not previously productive ($p=0.21$, numbers not provided in the paper). In a multiple regression model to predict publishing during residency (unclear whether publishing was a dichotomous variable representing any publication, or a continuous variable represent the number of publications), neither having published papers as a medical student ($p=0.91$) nor having published papers as an undergraduate ($p=0.11$) were statistically significant.

Robertson et al[17] performed a retrospective study of general surgery residents at a single academic medical center. The majority of the residents performed a 1-3 year research fellowship during their residency. The study was primarily examining the success that graduates from the program had on obtaining research funding after residency. They found that those who went on to receive any type of research funding published more papers during residency than those who did not, and specifically that those who went on to receive NIH research funding published more papers during residency than those who did not. However, it appears that the majority of people who applied for funding received it, and thus the comparison is likely confounded by former residents who are simply not interested in research.

Hellenthal et al[18] conducted a cross-section study of current and recently graduated chief residents in urology at US medical centers. They were primarily trying to characterize the publishing levels of these chief residents, but they found using a Poisson multivariate regression analysis that number of publications before residency was predictive of the number of manuscripts submitted for publication during residency ($IRR=1.11$, $p<0.001$).

Gill et al[19] conducted a cross-sectional study of internal medicine residents at a single Canadian Medical Center. They were primarily evaluating barriers to and predictors of publication of research by residents. They found in multivariate analysis that those with previous research experience were more lily to publish during residency (RR 1.6, CI 1.0-2.5).

Fang et al[20] conducted retrospective study of the participants of two Howard Hughes Medical Institute research scholarships for medical students: the HHMI Cloisters Program and the HHMI Medical Fellows Program. Each scholarship is for one year of research training. They were primarily examining career outcomes of program participants compared to non-awardees and also to MD-PhD graduates from the same time period. They found that awardees to the Cloisters Program (21% vs. 13%) and the Medical Fellows Program (24% vs. 10%) were more likely to receive NIH postdoctoral awards than non-awardees.

Ledley et al[21] conducted a retrospective study of those who completed a pediatric residency at a single institution from 1975-1981. They were primarily examining whether graduates from an academic residency were able to have successful academic careers. They conducted a MEDLINE search of all 150 graduates to ascertain their publication record, and found that there was no correlation between total citations before completion of residency and subsequent publication activity ($R2 = 0.14$). However, they also found that those who had at least one citation before completion of residency were significantly more likely to publish again after residency (96% vs. 62%, p value not reported) and had more citations (numbers not reported).

Association between Early Publishing and Future Success:

There is a significant body of literature on the issue of association between early publishing and future success. A systematic review (see above) uncovered fifteen articles that addressed the topic. The quality of the studies was typically quite low, as they examined the association either within a single specialty, a single institution, or both. Several of the studies were looking at predictors of future academic success, or at factors associated with choosing a career in academic medicine, and were not primarily designed to address the topic of early research or publishing.

The literature does demonstrate a fairly clear association between early research or publishing and future success. The majority of the identified articles demonstrated a positive association in at least one of the outcome measures, and several of the articles demonstrated the association specifically among academic physicians.

Research and Propensity to Enter an Academic Career:

There is a body of literature that demonstrates an association between doing research during medical school or residency and choosing to enter an academic career. A systematic review by Straus, et al.[3] examined the association of various factors and their association with deciding to enter a career in academic medicine. The review included several studies that showed a positive association between research experience during medical school and choosing to enter an academic career.[20, 22-26]

This association makes sense for two reasons. First, those who were already interested in careers in academia are probably much more likely than those who are interested in a career in private practice to do research in medical school (self-selection component). Second, students who were not initially interested in a career in academia but who chose to do research during medical school for some other reason (thought it would make them a better doctor, thought it would help them get into a better residency, etc.) are more likely to choose a career in academia since they have been exposed to it (true effect component).

While this association represents a potential benefit of medical student research, it also complicates the study of effect of early research on future academic success. Cohort studies that compare those who did research and those who did not do research during medical school are likely to have unequal proportions who entered an academic career. Furthermore, all positions in academic medicine are not equal. In the recent past, there have developed two separate tracks within academia: the clinician-investigator (CI) and the clinician-educator (CE).[27, 28] Historically, promotions within academic medicine have been largely based on number of publications.[29] While studies have shown that CEs are not promoted as quickly as CIs[30], they are not expected to publish as many papers to be promoted.[28] While it has not been specifically addressed to my knowledge, it is possible that in addition to research experience in medical school being associated with choosing a career in academic medicine in general, it may also be associated with specifically becoming a CE. Thus, when measuring the effect that early research has on academic output, it may not be enough to say that the cohort is composed of medical academicians without considering more specifically how the cohort is composed. In some cases, it may be more appropriate to examine the effects on CIs and CEs separately.

Evaluative Bibliometrics and Citation Analysis:

Evaluative bibliometrics is a field of Information Sciences whose aim is to construct tools to quantitatively evaluate research performance.[31] Citation analysis is a tool within the field of evaluative bibliometrics which the number of times that a certain paper has been cited in the footnotes or bibliography of other papers is calculated, and used in other metrics as a marker for quality of the paper. Being able to objectively research quality is important for many reasons, but is a particularly important tool for government and funding agencies to prioritize research funds to the scientist, research team, or institution that has the greatest chance of a positive impact. Those same measures and tools, however, can also be used to measure research quality as an outcome and allow for the study of what factors are associated with high research quality.

The Web of Science (WoS) is a citation database tool offered by Thomson-Reuters. It consists of three databases: Science Citation Index Expanded, Social Sciences Citation Index, and Arts & Humanities Citation Index.[32] The Science Citation Index Expanded covers over 7100 journals,[32] and has excellent coverage of clinical medicine.[31] Scopus and Google Scholar are newer citation databases. Studies have been done to compare the three databases, but as all three are continually expanding, such research is quickly obsolete.[31] WoS, Scopus, or Google Scholar could have been used in the current study. However WoS is most commonly used in citation analysis, and both Duke and UNC have subscriptions to the WoS database, and neither have subscriptions to Scopus. The interface and search methods on WoS seemed to be better than on Google Scholar. These factors are why we decided to use WoS in our searches.

Another tool used in citation analysis is the Journal Impact Factor (JIF). The JIF is a way of measuring entire journals by their impact in their given field.[33] It is calculated by dividing the number of times that articles in that journal have been cited by the total number of significant papers published that year.[33] Articles such as letters, commentaries are not counted in the denominator, but can contribute (usually insignificantly to the numerator.[33] Thomson Scientific offers another similar measure, called the Journal Performance Indicators (JPI).[32] It is calculated the same way as the JIF (citations divided by total articles), but by using essentially the same method as the WoS and linking each individual article to the articles that cited it, it is able to calculate excluding editorials, letters, and other less significant article types.[33]

The JIF and JPI offer a way to evaluate entire journals, which serves important functions, such as helping libraries select which journals to subscribe to and helping authors decide which journals to submit manuscripts to.[33] (Ironically, the JIF is also used by the WoS to decide which journals to cover in its databases.[31]) However, the role of the JIF and JPI in citation analysis of individual scientists is less clear. Before online citation databases, one method of evaluating the bibliography of a researcher was to multiply each article by the JIF of the journal that it appeared in. However, since we now have easy direct access to the number of citations of individual articles, it may make more sense to use that in evaluating a researcher's body of work. For example, it is entirely possible that an article in a more prestigious journal could have significantly fewer citations (and subsequently a lower impact) than an article published in a less prestigious journal. In that case, the higher impact article in the less prestigious journal should be considered a higher quality scientific contribution. One advantage that JIF and JPI do have over evaluating articles based on number of citations is that it takes time for the article to "mature" and obtain citations, whereas the JIF and JPI provide a metric that can be used to evaluate articles before they have even been published.

A relatively new way to measure scientific output is the h-index, named after Dr. Hirsch, who introduced it.[34] The h-index equals h, where an author has h articles that have been cited at least h times. Thus, an h-index of ten would mean that an author has published ten papers that been cited at least ten times each. In his seminal paper, Hirsch argues that the h-index is a better single measure than either total publications or total citations. Total publications is a good measure of productivity, but it tells nothing about the quality or impact of the papers, and Hirsch argues that total citations can also be inflated with several "big hits."[34] Another advantage of the h-index is that it should increase linearly over time for a researcher who is publishing a stable rate of similar quality research, such that

$$h \sim mn, \qquad [1]$$

where m is the slope of the increase of h over time, and n is the number of years.[34] The factor m can then be used to compare researchers at different career stages.

The h-index is meant to be a simple index to grade both the quality and quantity of a researcher's contribution. However, there are shortcomings in the index that are worth mentioning. Disadvantages of the h-index are that it is not a good way to compare researchers across different fields. It also severely punishes researchers who publish only a few very high quality studies (references?).

Another problem not addressed by any of these measures is the issue of co-authorship (which is beyond the scope of this paper, but worth at least mentioning). In measures such as total publications, total citations, or h-index, no weight is given to the level of contribution of the author (co-author vs. first author vs. senior author). A small contribution made to a "big hit" publication can inflate the total number of citations. For example, during the course of our study, we came across an article[35] with over 500 authors that had over 7500 citations! Each of those authors, regardless of their contribution to that study, would have greater than twice the number of citations as the average researcher in our study! Over time, small contributions made to many papers can also inflate the publication count. There have been solutions proposed to the problem of co-authorship[36], but it continues to be an issue in all of academia.

"Author name disambiguation" is the term used to describe attributing a work of literature to a specific person. It is important in evaluating the body of work of scientists. Through our searches of the literature, we were unable to find a detailed description of a practical author name disambiguation strategy that could be done manually on a relatively few number of authors. Much of the published literature on citation analysis comes from the field of information science, and is more concerned with automated processes that can be used on large data sets.[7]

<u>Addendum to Methods and Materials</u> (See page 4):

The analysis of the four main outcomes involved a formal examination of potential confounders. We initially examined the characteristics of the sample with univariate analysis to assess the distribution of the variables. Means for each variable in the total population and by main exposure type were tabulated (see Table 2 in the manuscript). Next, we used bivariate analysis to examine the association of confounders with each of the four main outcomes. We used Pearson's correlation analysis for the only continuous variable (years since medical school graduation), and used t-tests or one-way analysis of variance for categorical variables.

In addition to the outcomes examined in the manuscript, we also examined the association of early publishing on h-factor (calculated only on papers published after medical school graduation through the end of 2007). We constructed a multiple linear regression analysis of covariance (ANCOVA) model. Like the analyses for the publications and citations post-medical school, the model was adjusted for years since medical school graduation (as of the end of 2007), gender, whether the participant went to medical school in the US or not, and presence of an additional graduate level degree.

Since the citation count acquired on WoS continually increases as papers get cited, we felt that it was important to do the entire analysis in a short time window to avoid introducing any measurement bias. All of the WoS searches were conducted between May 30 and June 8 in 2009. While citation counts for the later searches probably increased slightly over this time period, we feel that it was minimal enough to be considered negligible.

Justification for Exclusion of Researchers from Main Analyses:

Of the 102 researchers who responded to our CV request, we excluded a total of 19. Reasons for exclusion included researcher had a PhD (6), CV did not contain a publication list (4), publications on CV were marked as "selected publications" (3), CV was not updated through the end of 2007 (2), insufficient information on CV (3), and CV was sent as an un-openable file (1).

We excluded those who had a PhD. We were surprised that six people actually had a PhD that was not listed with their name in the article, but since our study was looking at "MD only" researchers, we felt it was appropriate to exclude these people.

We excluded those without a publication list and those marked as selected publications because it was critical to know the year of the first publication. We could have done a PubMed search to get a reasonable publication list, but there would have been a very high probability of misclassifying the time of first publication. Therefore, we felt that it was necessary to exclude these people.

Similarly, we also needed to know the year of medical school graduation to correctly classify the time of first publication. Our author disambiguation strategy also required us to know what institutions the researcher had been associated with (we did not require that this be pristine, but we did need to at least know roughly where they had been). Three CVs did not contain the year of graduation from medical school, or any indication on where the person had been during their career, and so we felt it necessary to exclude these people.

Perhaps the least compelling reason for exclusion was that the CV was not updated through the end of 2007. Since we identified each of the researchers because of a high impact article published in 2008, we fully expected each researcher's CV to be updated at least through 2007. We probably could have done a PubMed search to get a rough idea of the number of publications for those years, but since we used the number of publications from the CV as the "gold standard" for everyone else, we thought that it was better to exclude these two rather than create a new and arbitrary algorithm after-the-fact for these two.

Potential Confounders:

Since we had access to the full CV of each researcher, we could potentially examine a number of potential confounders. However, we wanted only to consider factors that were likely to affect the outcome of academic success, and that could easily and accurately be ascertained from each of the CVs. The confounders that we did decide to include were sex, time since medical school graduation, whether they attended medical school in the US or elsewhere, and whether they had another graduate level degree.

We did not ask the researchers to self report gender. We thought that to ask in our email for the researchers to tell us their sex when they sent their CV would have either been largely ignored, or have been interpreted as insulting, and may have lowered our participation rate. We decided that it would be possible to ascertain the sex on our own. In most cases, it was very easy to determine the sex from the name and context from the CV. For example, many included spouses name on their CV. In cases where the name was more ambiguous and no contextual clues were present, we aired on the side of caution and searched the web for a profile picture. Most of the researchers in our group held academic positions, and the institutional website often had profiles of our researchers, including a picture. There were no cases in which we attempted to find a picture that we were unable to do so.

One potentially important confounder that we were not able to control for was residency type (or practice specialty type). Our hypothesis is that typical career outputs differ significantly between the top researchers in the different fields. We extracted residency type from the CVs, but to our surprise, more than half of our researchers were in a subspecialty of internal medicine. Surgery and Ob-Gyn had only 5 each. Therefore, we felt that we did not have enough data points to include.

When sending out our initial email, we did ask the researchers to report year of birth if it was not present on the CV. We had hoped to examine whether age at graduation had any influence either on whether the researcher published during medical school (exposure), or on subsequent publication success (outcome). It may not have ended up being a very good potential confounder, since most people go to medical school at about the same age. However, we had hypothesized that early publishers may have been younger than late publishers. The reason we hypothesized this is that doing research in medical school could potentially delay graduation, and it may be that younger medical students would be more likely to delay graduation than someone who started medical school later, and may have taken some time off before medical school. However, it became obvious very quickly that a significant number of people who did not have their year of birth on their CV did not actually include it in the reply email to us. In hindsight, it probably would have been easier to not ask for it in the first place, since it was not likely to add much to our analysis.

Initially, we had hoped to identify the year of graduation from residency. It may be that the end of clinical training would be a more appropriate time point to "start the clock" on the beginning of an academic career. However, we quickly realized that it was very difficult to determine when someone graduates from residency. While it may be more typical for a physician to complete a "residency," and in some cases a clinical "fellowship" immediately afterwards (which we would have considered to be clinical training, the same as residency), we found that many people took time off before completing a fellowship. Furthermore, it was often difficult to distinguish between a clinical fellowship and a research fellowship. Many fellowships incorporate a significant amount of protected research time, and so it may not be possibe to fully differentiate the two anyway (or accurate to even try).

Addendum to Results (See page 7):

In association of confounders with the main exposure (Table 2), it appeared that time since medical school graduation appeared to be the most significant difference between the groups. Early publishers had graduated from medical school 4 years later than middle publishers and 5 years later than late publishers, on average. Since time since medical school graduation was strongly associated with more publications and citations, controlling for that factor increased the differences in both total publications and total citations in our adjusted models. Gender also appeared to be unevenly distributed among the groups, with women being much less likely to be early publishers. Since being a woman was associated with having fewer publications, controlling for that factor had the opposite effect of time, and attenuated the difference between early publishers and the other two groups.

In bivariate associations between the potential confounders and the outcome measures (Tables 5-9), it appeared that whether the researcher went to medical school in the US was least associated with differences in the outcomes, as it was not statistically significant in any of the analyses. Although not statistically significant, it was still somewhat surprising that having gone to medical school in the US was associated with lower outcome measures in each of the analyses. This may represent somewhat of a bias against foreign researchers, in that only the highest quality research from the best foreign researchers gets into the most prestigious US medical journals. Residency type was also non-significant in each of the analyses, but this may be because of the few data point for several of the residency types. Each of the other factors was significant in some of the analyses.

Time since medical school graduation was the only continuous variable, and it only appeared in the models for total publications, total citations, and h-index. In all three models models, test for linear association was highly significant ($p<0.001$). This was expected, as total publications, total citations, and h-index can only increase with time, and should increase relatively consistently over the course of an academic career.

Our examination of Pearson's correlation among nominal variables showed that there was no significant overlap between variables (data not shown). The strongest correlation was between career stage and early publishing status ($r=0.304$), which was well below a typical cut-off of $r=0.4$ or $r=0.5$. This strong association was likely due to the fact that nearly all of those who were 0-8 years since medical school graduation were early publishers, and none of them were late publishers. This association also makes sense, since to have published in NEJM or JAMA at such an early stage in their career, these researchers were very likely to have started publishing either during medical school or early in residency.

We used analysis of covariance (ANCOVA) regression models for each of the four main outcomes. We initially included all potential covariates (except for residency type) in the model. The model was then reduced by sequentially removing variables that were not confounders of the exposure/outcome relationship (i.e. did not significantly change the number of publications or citations estimates for each category of time of first publication). It appeared that for most outcome models, time since medical school graduation (career stage in two of the models) was the most important confounder, and that gender and the location of medical school were also weak confounders. Whether the researchers had an additional masters degree was not a confounder in any of our models, as the means for publications and citations using a model this variable present were nearly identical to the means of the full model. In the end, we decided to leave all of the variables in the model as we felt them to be potentially important confounders.

The result of our analysis of h-index is shown in Table 10. Similar to the results of the publication and citation analysis, the h-index showed a slight increase for early publishers in the unadjusted analysis. The difference was increased and became statistically significant on adjusted analysis.

To examine the effectiveness and potential bias of our WoS search, we measured the total number of publications identified for each of the groups in both the post medical school and the 06-07 searches (Tables 11-12). As expected, the ratio for all groups was between 60% and 80%. This is because some of the publications, such as book chapters and publications in smaller journals, are not included in the WoS database. As indicated, WoS contained a slightly higher proportion of the Early Publishers' publications, and a roughly equal proportion of the Middle and Late Publishers' publications for both time periods. It makes some sense that WoS would contain a higher proportion of the total publications for a younger researcher, because older papers may be either less likely to be in the database or coded in such a way that was more difficult to find using our search strategy (some older papers did not include institutional affiliation). However, for the 2006-2007 searches, such an explanation does not exist, and it is unclear why the ratios would differ. However, in both cases, the difference was minimal, and unlikely to significantly affect our findings.

<u>Addendum to Discussion</u> (See page 9):

The history of medical student research is well documented. Charles Best was a medical student when he was involved in the discovery of insulin,[37] and Niels Stensen, Johann Friedrich Meckel, and Paul Langerhans were all students when each discovered the anatomical structures that bear their names today.[1] However, as modern medicine has moved into the era of "evidence-based medicine," medical education also needs to move into a time when anecdotes of success are replaced by rigorous research that proves that the methods used to train future physicians and physician-scientists work. Research is an important component in the education of both physicians and physician-scientists, but many questions are left unanswered.

The first question that needs to be addressed is whether research during medical school is beneficial for all medical students. One of the benefits of research is that it is associated with entering a career in academic medicine.[3] Biomedical research as a whole is critically important to the well-being of our society, and medical school may be an excellent time to identify who has potential to be a great researcher. The more students who want to go into research, the more competitive the positions will become, which should improve the quality of research. If that is the motive for having students participate in research, however, studies need to be conducted on the length of research that is most beneficial. It may not be necessary to have all medical students complete a year of research if the goal is "whet their appetite," in which case a few months may suffice. Further, the effectiveness of medical student research in causing medical students to enter academic careers needs to be measured against other interventions that accomplish the same goal, such as reducing medical student debt.[3]

Another reason that has been stated for having all medical students participate in research is that it may help to "instill a culture of evidenced-based medicine"[37] and "sharpens his judgment and enhances his evaluation of the research of others."[2] This may be true, but to my knowledge it has not been studied. Additionally, other methods may also instill a culture of evidence-based medicine, such as formal courses in clinical epidemiology and critical appraisal of the literature. Such interventions should be compared in their ability teach physicians to apply new research to their practices.

Another important question is, "when is the optimal time for a physician-scientist to begin his or her research career?" Victor Johnson said in 1959, "the methods and tools of research and the nature and conduct of investigation must be learned early."[2] This is undoubtedly true, but how early? Many residencies and fellowships have research time built into their curriculum, whereas medical students may need to delay their graduation by a year to undertake a significant research project during medical school. Furthermore, a resident or fellow may have more insight and expertise on their research topic, which allows them to take a more prominent role in the design of their project. Medical students often work with a mentor who is an expert on their topic, and while it may be beneficial simply to participate in the conduction of a mentor's project, it is unlikely that a medical student will be able to contribute any insight or expertise in the design of the project. Studies need to be done to determine whether research that is delayed until residency or fellowship is as beneficial as research conducted during medical school.

Once it has been decided that research during medical school is beneficial, the next step is to determine exactly what that research should look like. At some medical schools, a typical research project is several months long.[12] At Duke[38] and Stanford[39], and in many of the research grants available for students, such as the Doris Duke Fellowship[40], the Sarnoff Fellowship[40], and the HHMI Cloister Program and HHMI Clinical Fellows Program[20], the research is approximately one year. Additionally, what role should the student play in the design of the research. If the goal of the research is simply to give a student a hands-on interaction with a basic science experiment, or hands-on experience collecting and analyzing clinical data, then it makes more sense for the student to be paired up with a mentor who will be very instrumental in the design of the project. If, however, the goal is to allow students the opportunity to think critically about a problem, design an experiment to answer the question, and then interpret their findings, it makes more sense for the mentor to be "hands-off," and only help students when they run into a problem. However, this latter design is unlikely to be accomplished in a short, 3-month project. Furthermore, for students to design a research project, it requires that they have some expertise in the field, and some insight into a gap in the current literature, which is something that students often do not have. If this is the type of research that is deemed to be the most beneficial, then perhaps it would make more sense to delay research until residency.

However, studies to date have not conclusively shown that research during medical school effects future academic success. Several of the studies identified in our systematic review display serious design flaws, at least for answering the question of true effect. Any study comparing the careers of successful and unsuccessful applicants to a research scholarship will not be able to determine the effect of the research scholarship, because one would expect that those who were awarded the scholarship would be better than unsuccessful applicants in ways that would be difficult to control for. Studies in which students or residents who did research or published are compared to their peers who did not are also problematic. As long as research is optional, students who are interested in research will self-select to do it. Those students are likely to be different in ways that are impossible to control for, such as higher scientific curiosity or motivation. The best designed study was by Segal, et al[11] in which students from three similar medical schools (one requiring research, and two not requiring research) were compared, did not adequately characterize the future academic careers of the graduates from the three schools. However, in the measures that they did make, including percentage with academic appointments, there did not appear to be any differences between the graduates from the three schools.

To our knowledge, there have been no studies examining the negative aspects of medical student research. While the whole of medical training is very long, medical school itself is relatively short. Thanks to the recent advances in genetics and molecular biology, the amount medical information has increased rapidly in the recent past, and therefore the potential amount of information learned by medical students continues to expand. Additionally, there continues to be pressure to increase the scope of the curriculum, such as formally teaching ethics.[41] It is possible that there is simply too much to learn in four years for medical students to take time away from their studies. Research needs to be done to determine whether there are any negative impacts of medical student research.

While this study by no means proves that research during medical schools leads to improved academic careers, it does raise the possibility of such an effect. To our knowledge, this is the first study that has shown that even among research intensive academicians, early publication is associated with future publication success. Such a strong association is necessary in order for there to be a causal relationship, as pointed out by Hill.[42] However, Hill also said that plausibility was an important component to proving causality. Our initial hypothesis was that whatever effect that early publishing would have would decrease over time, and that those who began their research careers later would eventually "catch up," but this is not what we found. The effect was strong, and it did not appear to attenuate over time. It is certainly possible that effect was simply stronger than we hypothesized, but it remains possible that the results of our 2006-2007 analysis are due primarily to some bias in our study that we were unable to account for, such as some personal characteristics that made the early publishers naturally better researchers.

Limitations:

It is very clear to us now that citation analysis is a very inexact science. There is no doubt that in our analysis, some papers were missed, and some papers were incorrectly attributed to our researchers. However, we feel that while the results should be evaluated with some degree of skepticism, the results do have some significant meaning. Additionally, the fact that results of our citation analyses agreed so well with the results of our publication analyses.

As the field of author disambiguation progresses, there will likely be tools in the near future that makes this type of analysis more accurate and easier to do.

Future Studies:

Having found that the association between early publishing and future publishing success holds in a selected group of research-intensive academicians, the next step is to further explore the possibility of a causal relationship. This line of research will need to deal with potential confounders, such as personal characteristics like motivation to do research or enter a research career. The gold standard in biomedical research for dealing with confounding of this nature is to perform a randomized controlled trial. However, an RCT in this case is not possible, and thus other methods for dealing with confounding will have to be considered.

One possibility would be to use propensity scores to control for confounding. Unlike conventional linear regression, which adjusts the outcome based on factors that affect the outcome, propensity scores are used to adjust the outcome based on factors that affect the exposure. In this way, cohorts could be constructed that had the same propensity to do research in medical school, and then the outcomes of the cohorts with equal propensity scores could be compared to give an unbiased result. A propensity score in this case could be constructed using factor such as where the researcher went to medical school, where they went to undergraduate, gender, age at matriculation, occupation of their parents, and other factors that would affect the propensity of the student to do research.

Another option for dealing with the potential confounding of personal characteristics would be to measure those characteristics prospectively. For example, an incoming cohort of medical students could be given Myers-Briggs tests, and other surveys to measure their desire to do research. Grades in preclinical and clinical courses in medical school could then be used as a measure of how hard working the student was. Adding all of these "early confounders" to "late confounders", such as residency type, could then be used to construct a linear regression model to give an unbiased (or at least less biased) result of effect of early research or publication. However, any prospective method would require a very long time for the research careers to mature, and a large sample size would be required to detect what is likely a very small effect would make such a study prohibitively expensive.

Addendum to Figures and Tables (See page 11):

Table 4. Systematic Review Evidence Table

Author, Year	Participants	Exposure	Outcome	Effect Size	p value	Quality Rating Internal Validity	 External Validity
Brancati, 1992	424	Research Experience in Med School	Higher attained faculty rank	RR=3.11	0.0001	good	fair
	424	Research Experience in Med School	Citations to published work	ratio=2.42	0.0008		
Dorsey, 2006	68	Number of publications before residency	Post-residcency publication score	R^2=0.11	0.01	good	fair
	68	Number of publications during residency	Post-residcency publication score	R^2=0.26	<0.001		
Evered, 1987	54	1) Intercalated degree 2) Oxford or Cambridge 3) Other med school with no intercalated degree	citations per paper	1) 8.04 2) 7.63 3) 4.16	?	poor	poor
Reinders, 2005	274	Extracirricular research experience during medical school	Average number of publications after medical school	4 vs. 1	?	poor	poor
Segal, 1990	567	Research in medical school	Percent who published research after Med school	49% vs. 32%	<0.001	fair	fair
Chusid, 1993	79	Publication of medical school thesis	Publications during 20 years after medical school	22.1 vs. 14.4	0.0005	fair	poor
Dyrbye, 2008	981	Published a research report related to required research	Publications unrelated to required research during first 3 years after medical school	0.5 vs. 0.3	<0.0001	fair	fair
	998	Published an abstract related to required research	Publications unrelated to required research during first 3 years after medical school	0.5 vs. 0.4	0.04		
	920	Research presentation at a meeting	Publications unrelated to required research during first 3 years after medical school	0.5 vs. 0.3	0.001		

Table 4 Continued.

Author, Year	Participants	Exposure	Outcome	Effect Size	p value	Quality Rating	
						Internal Validity	External Validity
Lessin, 1995	200	Laboratory research experience during general surgery residency	Laboratory publications during pediatric surgery residency	0.8 vs. 0.4	>0.05	poor	poor
	200	Laboratory research experience during general surgery residency	Laboratory publications after pediatric surgery residency	1.5 vs. 1.4	>0.05		
	200	Laboratory research experience during general surgery residency	Clinical publications during pediatric surgery residency	2.7 vs. 2.8	>0.05		
	200	Laboratory research experience during general surgery residency	Clinical publications after pediatric surgery residency	5.2 vs 5.7	>0.05		
	200	Had at least 1 publication during general surgery residency	Number of publications during pediatric surgery residency	4.5 vs. 2.1	?		
	200	Had at least 1 publication during general surgery residency	Total number of publications after general surgery residency	11.9 vs. 5.1	?		
Smith, 2009	1159	Awarded APS-SPR MSRP research scholarship (compared to unsuccessful applicants)	Percent of applicants who had published	79% higher for awardees	<0.0005	fair	poor
	1159	Awarded APS-SPR MSRP research scholarship (compared to unsuccessful applicants)	Publications per applicant	higher for awardees (numbers cannot be determined from paper)	0.007		
Patterson, 2002	73	"Previously productive" during medical school	publications during residency and year following completion of residency	numbers cannot be determined from paper	0.21	fair	poor
Robertson, 2007	67	Number of publications during residency	Received any funding after residency	9.3 vs. 5.2	0.02	fair	poor
	67	Number of publications during residency	Received NIH funding after residency	10.2 vs. 5.9	0.03		
Hellenthal, 2009	127	Number of publications before residency	Number of manuscripts submitted during residency	IRR = 1.11	<0.001	fair	poor
Gill, 2001	81	Previous research experience	Publication during residency	RR=1.6	CI 1.0-2.5	poor	poor
Fang, 2003	867	HHMI Cloister Program Participant (vs. non-awardee)	Receipt of NIH postdoctoral award	21% vs. 13%	<0.010	good	poor
	364	HHMI Medical Fellows Program participant (vs. non-awardee)	Receipt of NIH postdoctoral award	24% vs. 10%	<0.001		
Ledley, 1992	150	Publication prior to completeion of residency	Publication after completion of residency	96% vs. 62%	?	fair	poor

Table 5. Bivariate Associations between Patient Characteristics and Total Publications after Graduation from Medical School

Variable	Publications: Mean or Correlation	p value
Publishing		
Early	137	0.229
Middle	119	
Late	77	
Gender		
male	132	0.013
female	62	
MS in US		
yes	108	0.349
no	132	
Years since grad	r=0.690	<0.001
Residency type		
IM	114	0.919
Surg	128	
Peds	131	
Ob-Gyn	74	
Other	118	
Other Degrees		
none	122	0.072
MPH	53	
Other Masters	136	

Table 6. Bivariate Associations nbetween Patient Characteristics and Total Citations on Publications after Graduation from Medical School

Variable	Publications: Mean or Correlation	p value
Publishing		0.234
Early	3925	
Middle	3177	
Late	1919	
Gender		0.003
male	3806	
female	1024	
MS in US		0.511
yes	2953	
none	3535	
Years since grad	r=0.686	<0.001
Residency type		0.800
IM	3387	
Surg	2914	
Peds	3281	
Ob-Gyn	1156	
Other	2995	
Other Degrees		0.110
none	3710	
MPH	1269	
Other Masters	3046	

Table 7. Bivariate Associations Between Patient Characteristics and Publications in 2006-2007

Variable	Publications: Mean or Correlation	p value
Publishing		0.168
Early	25.0	
Middle	19.9	
Late	15.3	
Gender		0.013
Male	22.9	
Female	12.6	
MS in US		0.078
yes	18.3	
no	25.2	
Career Stage		0.010
0-8	6.0	
8-16	17.7	
16-24	21.5	
24+	26.0	
Residency type		0.974
IM	21.4	
Surg	17.6	
Peds	20.3	
Ob-Gyn	20.6	
Other	18.6	
Other Degrees		0.531
none	20.5	
MPH	16.2	
Other Masters	22.6	

Table 8. Bivariate Associations between Patient Characteristics and Citations on Publications in 2006-2007

Variable	Publications: Mean or Correlation	p value
Publishing		0.302
Early	298.5	
Middle	195.1	
Late	160.6	
Gender		0.060
Male	254.0	
Female	104.5	
MS in US		0.436
yes	199.9	
no	257.6	
Career Stage		0.082
0-8	49.4	
8-16	146.0	
16-24	230.3	
24+	312.2	
Residency type		0.434
IM	271.3	
Surg	65.8	
Peds	180.8	
Ob-Gyn	205.2	
Other	136.0	
Other Degrees		0.771
none	240.2	
MPH	194.2	
Other Masters	188.3	

Table 9. Bivariate Associations between Patient Characteristics and h-index on Publications after Graduation from Medical School

Variable	Publications: Mean or Correlation	p value
Publishing		0.212
Early	25.6	
Middle	24.1	
Late	16.9	
Gender		0.001
male	26.2	
female	13.1	
MS in US		0.308
yes	21.8	
no	25.8	
Years since grad	r=0.709	<0.001
Residency type		0.835
IM	23.6	
Surg	23.0	
Peds	26.4	
Ob-Gyn	16.2	
Other	21.5	
Other Degrees		0.109
none	25.0	
MPH	14.4	
Other Masters	23.9	

Table 10. Results of Analysis of h-index

Results	Early Publishers Mean (n = 24)	Middle Publishers Mean (n=42)	Late Publishers Mean (n = 17)	p value
Unadjusted H-index on publications after medical school	25.6	24.1	16.9	0.2119
Adjusted H-index on publications after medical school	28.9	23.0	15.0	0.001ǂ

* Adjusted for years since med school graduation, gender, location of med school (US or not US), and additional degrees (none, MPH, or other masters)

ǂ Overall p value (2df); For individual comparisons: Early vs. Middle, p=0.039; Early vs. Late, p<0.001; Middle vs. Late, p=0.010.

Table 11. Ratio of Publications on WoS to Publications on CV after Medical School Graduation

	WoS Publications mean	CV Publications mean	RATIO
Early Publishers	96.33	136.88	0.704
Middle Publishers	78.81	118.57	0.665
Late Publishers	50.12	77.29	0.648

Table 12. Ratio of Publications on WoS to Publications on CV for 2006-2007

	WoS Publications mean	CV Publications mean	RATIO
Early Publishers	19.54	25.04	0.780
Middle Publishers	14.12	19.88	0.710
Late Publishers	10.88	15.29	0.712

Acknowledgements:

I owe many thanks to the following people:

- Zachary Reitman, for helping me come up with the idea for this study, for helping me send the emails, and for the many, many hours spent pouring over the CVs and on Web of Science with me.
- Dr. Thelma Mielenz, PhD, for her editing help, statistical guidance, and encouragement.
- Dr. Philip Goodman, MD, for his guidance in planning this project and his editing help.
- Dr. Russell Harris, MD, for his guidance throughout the year, and working with me and allowing me to change my paper topic so late in the year.

Appendix:

Table 13. Systematic Review Literature Searches

Date	Database	Main search terms	Modifiers	Yield	Used search?
6/16/09	Medline	(medical students OR medical education OR research experience" AND (academic success OR academic medicine OR medical academia OR publication success)	none	11097	no
6/24/09	Medline	"education, medical" [MeSH Terms]	1970-current	91708	no
6/24/09	Medline	"publishing" [MeSH Terms]	1970-current	26385	no
6/24/09	Medline	"Research Support as Topic" [MeSH Terms]	1970-current	16741	no
6/24/09	Medline	"Educational Measurement" [MeSH Terms]	1970-current	77586	no
6/24/09	Medline	"research" [MeSH Terms]	1970-current	162152	no
6/24/09	Medline	"achievement" [MeSH Terms]	1970-current	9095	no
6/24/09	Medline	"career mobility" [MeSH Terms]	1970-current	7887	no
6/24/09	Medline	"research personnel" [MeSH Terms]	1970-current	8760	no
6/24/09	Medline	("education, medical" [MeSH Terms] OR "Educational Measurement" [MeSH Terms]) AND ("publishing" [MeSH Terms] OR "Research Support as Topic" [MeSH Terms] OR "achievement" [MeSH Terms] OR "career mobility" [MeSH Terms])	1970-current	4693	no
6/24/09	Medline	"education, medical"[MeSH Terms] AND ("publishing"[MeSH Terms] OR "Research Support as Topic"[MeSH Terms] OR "achievement"[MeSH Terms] OR "career mobility"[MeSH Terms])	1970-current	2131	no

6/24/09	Medline	"education, medical"[MeSH Terms] AND ("publishing"[MeSH Terms] OR "Research Support as Topic"[MeSH Terms] OR "achievement"[MeSH Terms] OR "career mobility"[MeSH Terms])	1970-current AND English	1906	no
6/24/09	Medline	"education, medical"[MeSH Terms] AND ("publishing"[MeSH Terms] OR "Research Support as Topic"[MeSH Terms] OR "career mobility"[MeSH Terms])	1970-current AND English	1491	no
6/24/09	Medline	"education, medical"[MeSH Terms] AND ("publishing"[MeSH Terms] OR "Research Support as Topic"[MeSH Terms] OR "achievement"[MeSH Terms] OR "career mobility"[MeSH Terms])	1975-current AND English	1674	yes

Figure 1. Initial Email

Subject: Duke med student research project

Dr. X,

I am a third year student at Duke University School of Medicine. I obtained your email address from your recent NEJM article. I am writing on behalf of myself, another third year medical student, and our faculty mentor to ask your assistance on a research project. We are conducting a study to look at the relationship between time of first scientific publication and future academic success, specifically future publication rates and sequence of academic promotion. Our study involves extracting information from the CVs of successful academicians, and an online search using Web of Science. We are writing to ask for your help in our investigation. If you decide to help, please email us your most recent CV, and your year of birth if not contained in your CV. It is important that your CV contain all of your publications, especially your earliest ones.

Feel free to remove any sensitive information from your CV, such as social security number, phone numbers, addresses, etc. We will only use your information for the current study, and will guard your privacy and confidentiality to the highest standard. Our study has been reviewed and exempted by the Duke IRB. If you have any questions, please contact me.

Thank you in advance for your assistance in this endeavor.

Kevin Riggs
Duke University School of Medicine

Figure 2. Follow-up Email

Subject: Duke med student research project

Dr. X,

I am writing to follow up the email I sent you two weeks ago. In short, I am conducting a study to explore the association between early publication and future academic success among medical researchers. I am collaborating with another third year medical student and Philip Goodman M.D., Associate Dean for Medical Education on this project. Please consider sending a current CV and your year of birth to aid in this study.

Thank you in advance for your consideration.

Kevin Riggs
Duke University School of Medicine
MS3

References:

1. Hendrix M. The legacy of medical students in medical research. *JAMA.* Dec 6 1995;274(21):1735, 1739.

2. Johnson V. Value of research for the graduate and post-graduate medical student. *J Am Med Assoc.* Sep 5 1959;171:24-26.

3. Straus SE, Straus C, Tzanetos K. Career choice in academic medicine: systematic review. *J Gen Intern Med.* Dec 2006;21(12):1222-1229.

4. Wyngaarden JB. The clinical investigator as an endangered species. *N Engl J Med.* Dec 6 1979;301(23):1254-1259.

5. Brancati FL, Mead LA, Levine DM, Martin D, Margolis S, Klag MJ. Early predictors of career achievement in academic medicine. *JAMA.* Mar 11 1992;267(10):1372-1376.

6. Dorsey ER, Raphael BA, Balcer LJ, Galetta SL. Predictors of future publication record and academic rank in a cohort of neurology residents. *Neurology.* Oct 24 2006;67(8):1335-1337.

7. Smalheiser NR, Torvik VI. Author Name Disambiguation. In: Cronin B, ed. *Annual Review of Information Science and Technology.* Vol 43. Medford, NJ: Information Today; 2009:287-313.

8. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med.* Apr 2001;20(3 Suppl):21-35.

9. Reinders JJ, Kropmans TJ, Cohen-Schotanus J. Extracurricular research experience of medical students and their scientific output after graduation. *Med Educ.* Feb 2005;39(2):237.

10. Evered D, Anderson J, Griggs P, Wakeford R. The correlates of research success. *British Medical Journal (Clinical research ed.).* 1987;295(6592):241.

11. Segal S, Lloyd T, Houts PS, Stillman PL, Jungas RL, Greer RB, 3rd. The association between students' research involvement in medical school and their postgraduate medical activities. *Acad Med.* Aug 1990;65(8):530-533.

12. Dyrbye LN, Davidson LW, Cook DA. Publications and presentations resulting from required research by students at Mayo Medical School, 1976-2003. *Acad Med.* Jun 2008;83(6):604-610.

13. Chusid MJ, Havens PL, Coleman CN. Alpha omega alpha election and medical school thesis publication: relationship to subsequent publication rate over a twenty-year period. *Yale J Biol Med.* Mar-Apr 1993;66(2):67-73.

14. Lessin MS, Klein MD. Does research during general surgery residency correlate with academic pursuits after pediatric surgery residency? *J Pediatr Surg.* Sep 1995;30(9):1310-1313.

15. Smith WH, Rogers JG, Hansen TN, Smith CV. Early career development in academic pediatrics of participants in the APS-SPR Medical Student Research Program. *Pediatr Res.* Apr 2009;65(4):474-477.

16. Patterson SK, Fitzgerald JT, Boyse TD, Cohan RH. Is past academic productivity predictive of radiology resident academic productivity? *Acad Radiol.* Feb 2002;9(2):211-216.

17. Robertson CM, Klingensmith ME, Coopersmith CM. Long-term outcomes of performing a postdoctoral research fellowship during general surgery residency. *Ann Surg.* Apr 2007;245(4):516-523.

18. Hellenthal NJ, Ramirez ML, Yap SA, Kurzrock EA. Manuscript publication by urology residents and predictive factors. *J Urol.* Jan 2009;181(1):281-286; discussion 286-287.

19. Gill S, Levin A, Djurdjev O, Yoshida EM. Obstacles to residents' conducting research and predictors of publication. *Acad Med.* May 2001;76(5):477.

20. Fang D, Meyer RE. Effect of two Howard Hughes Medical Institute research training programs for medical students on the likelihood of pursuing research careers. *Acad Med.* Dec 2003;78(12):1271-1280.

21. Ledley FD, Lovejoy FH, Jr. Prospects for academically trained pediatricians in academic medicine. *Clin Invest Med.* Dec 1992;15(6):518-526.

22. Feng L, Ruzal-Shapiro C. Factors that influence radiologists' career choices. *Acad Radiol.* Jan 2003;10(1):45-51.

23. Hillman BJ, Fajardo LL, Witzke DB, Cardenas D, Irion M, Fulginiti JV. Influences affecting radiologists' choices of academic or private practice careers. *Radiology.* Feb 1990;174(2):561-564.

24. Neacy K, Stern SA, Kim HM, Dronen SC. Resident perception of academic skills training and impact on academic career choice. *Acad Emerg Med.* Dec 2000;7(12):1408-1415.

25. Sanders AB, Fulginiti JV, Witzke DB. Factors influencing resident career choices in emergency medicine. *Ann Emerg Med.* Jan 1992;21(1):47-52.

26. Bilbey JH, Fache JS, Burhenne HJ. Are there predictors for future academic radiologists? A Canadian survey. *Can Assoc Radiol J.* Oct 1992;43(5):369-373.

27. Greer DS. Faculty rewards for the generalist clinician-teacher. *J Gen Intern Med.* Jan-Feb 1990;5(1 Suppl):S53-58.

28. Beasley BW, Wright SM, Cofrancesco J, Jr., Babbott SF, Thomas PA, Bass EB. Promotion criteria for clinician-educators in the United States and Canada. A survey of promotion committee chairpersons. *JAMA.* Sep 3 1997;278(9):723-728.

29. Stossel TP. Volume: papers and academic promotion. *Ann Intern Med.* Jan 1987;106(1):146-149.

30. Beasley BW, Simon SD, Wright SM. A time to be promoted. The Prospective Study of Promotion in Academia (Prospective Study of Promotion in Academia). *J Gen Intern Med.* Feb 2006;21(2):123-129.

31. Moed HF. New developments in the use of citation analysis in research evaluation. *Arch Immunol Ther Exp (Warsz).* Jan-Feb 2009;57(1):13-18.

32. http://in-cites.com/rsg/jpi/. Accessed June 23, 2009.

33. Garfield E. The history and meaning of the journal impact factor. *JAMA.* Jan 4 2006;295(1):90-93.

34. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A.* Nov 15 2005;102(46):16569-16572.

35. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. *N Engl J Med.* Sep 30 1993;329(14):977-986.

36. Angell M. Publish or perish: a proposal. *Ann Intern Med.* Feb 1986;104(2):261-262.

37. Metcalfe D. Involving medical students in research. *J R Soc Med.* Mar 2008;101(3):102-103.

38. Grochowski C, Halperin E, Buckley E. A Curricular Model for the Training of Physician Scientists: The Evolution of the Duke University School of Medicine Curriculum. *Academic Medicine.* 2007;82(4):375.

39. Jacobs CD, Cross PC. The value of medical student research: the experience at Stanford University School of Medicine. *Med Educ.* Sep 1995;29(5):342-346.

40. http://www.sarnoffendowment.org/program/information.cfm. Accessed June 25, 2009.

41. Hafferty FW, Franks R. The hidden curriculum, ethics teaching, and the structure of medical education. *Acad Med.* Nov 1994;69(11):861-871.

42. Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med.* May 1965;58(5):295-300.